

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2003-50741

(P2003-50741A)

(43)公開日 平成15年2月21日(2003.2.21)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード(参考)
G 0 6 F 12/08	5 0 5	G 0 6 F 12/08	5 0 5 B 5 B 0 0 5
	5 0 3		5 0 3 Z 5 B 0 1 3
	5 1 3		5 1 3 5 B 0 3 3
	5 5 1		5 5 1 Z
	5 5 9		5 5 9 Z

審査請求 未請求 請求項の数24 O L (全 21 頁) 最終頁に続く

(21)出願番号 特願2001-239763(P2001-239763)

(22)出願日 平成13年8月7日(2001.8.7)

(71)出願人 000232036

エヌイーシーマイクロシステム株式会社  
神奈川県川崎市中原区小杉町1丁目403番  
53

(72)発明者 柴田 幸茂

神奈川県川崎市幸区塚越3丁目484番地  
日本電気マイコンテクノロジー株式会社内

(74)代理人 100082935

弁理士 京本 直樹 (外2名)

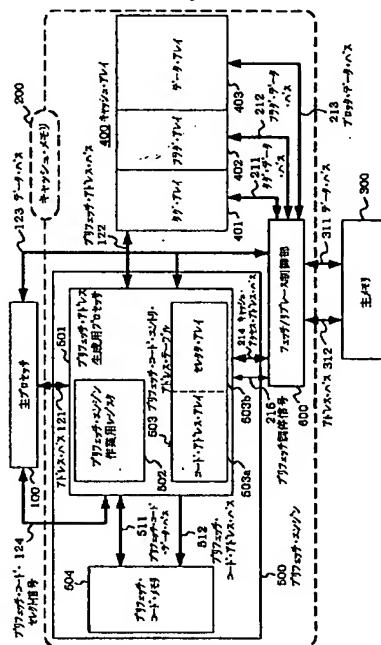
最終頁に続く

(54)【発明の名称】 ソフトウェア制御可能なキャッシュ・メモリ装置、制御方法およびプログラム

(57)【要約】

【課題】ソフトウェアで制御可能なプリフェッチ機構を実現する。

【解決手段】主プロセッサ100と主メモリ300間に設けられるキャッシュ・メモリ200が、プリフェッチ・アドレスの決定をプログラムの実行時に行うプリフェッチ手段として、実行すべきプリフェッチ・コードを指定する指示を前記主プロセッサ100から受けてプリフェッチ・アドレス生成をソフトウェアによって制御するプリフェッチ・アドレス生成用プロセッサ501を含むプリフェッチ・エンジン500をさらに備える。



## 【特許請求の範囲】

【請求項1】 主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリが、プログラム実行時に、プリフェッチ・アドレスをソフトウェア制御で決定するプリフェッチ手段として、実行すべきプリフェッチ・コードの指定を前記主プロセッサから指示されるとともに前記プリフェッチ・コードに対応した命令の実行により前記プリフェッチ・アドレスを生成するプリフェッチ・アドレス生成用プロセッサを含み、かつ前記主プロセッサと並列処理を行うプリフェッチ・エンジンを備えたことを特徴とするソフトウェア制御可能なキャッシュ・メモリ装置。

【請求項2】 前記専用プロセッサは、浮動小数点演算機能が省略された特定用途のプロセッサである請求項1記載のソフトウェア制御可能なキャッシュ・メモリ装置。

【請求項3】 前記専用プロセッサは、前記主プロセッサからの指示に従い、プリフェッチ・コード・メモリに格納されている専用プロセッサ自身の制御用命令を読み出し、読み出した前記命令を自律的に逐次実行して前記プリフェッチ・アドレスを生成するとともに、生成した前記プリフェッチ・アドレスを前記キャッシュ・メモリ内にあるフェッチ／リプレース制御手段に対して送出する機能を有する請求項1または2記載のソフトウェア制御可能なキャッシュ・メモリ装置。

【請求項4】 主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内に、アドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記専用プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ／リプレース制御手段と備えることを特徴とするソフトウェア制御可能なキャッシュ・メモリ装置。

【請求項5】 前記プリフェッチ・コード・メモリを設ける代わりに、主メモリにプリフェッチ・コードを格納し、主メモリを主プロセッサと専用プロセッサで共有する請求項3または4記載のソフトウェア制御可能なキャッシュ・メモリ装置。

【請求項6】 前記プリフェッチ・コード・メモリを設ける代わりに、前記キャッシュ・メモリ内に前記プリフェッチ・コードを格納する請求項3または4記載のソフトウェア制御可能なキャッシュ・メモリ装置。

【請求項7】 前記フェッチ／リプレース制御手段は、前記専用プロセッサから受け取ったプリフェッチ・アドレスに対応するキャッシュ・ブロックをフェッチして入れ替えるとともに、前記主プロセッサが次に実行するア

クセスまで待機する機能を有する請求項1または4記載のソフトウェア制御可能なキャッシュ・メモリ装置。

【請求項8】 前記プリフェッチ・エンジン作業用レジスタ・ファイルは、前記専用プロセッサがプリフェッチ・アドレスを生成するのに必要なデータを格納する機能を有し、前記専用プロセッサは前記主メモリを直接アクセスせずに前記プリフェッチ・エンジン作業用レジスタ・ファイルをアクセスする機能を有する請求項4記載のソフトウェア制御可能なキャッシュ・メモリ装置。

【請求項9】 前記主プロセッサは、前記プリフェッチ・コード・メモリに格納された、前記専用プロセッサが実行すべき命令列を指定するためのプリフェッチ指示コードをもつプリフェッチ・コード・セレクト信号を有し、そのプリフェッチ・コード・セレクト信号を前記専用プロセッサに転送する請求項3または4記載のソフトウェア制御可能なキャッシュ・メモリ装置。

【請求項10】 前記プリフェッチ・コード・セレクト信号のビット幅を、メモリアクセス命令にプリフェッチ・コードのアドレスを指定できるだけの情報に必要なビット幅に予め拡張して備える請求項9記載のソフトウェア制御可能なキャッシュ・メモリ装置。

【請求項11】 主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ／リプレース制御手段と備え、前記プリフェッチ・アドレスの決定をプログラム実行時に行うために、前記専用プロセッサが、前記プリフェッチ・コード・メモリに格納されている命令を読み出し逐次実行してプリフェッチ・アドレスを生成し、生成したプリフェッチ・アドレスを前記フェッチ／リプレース制御手段に対して送出し、前記フェッチ／リプレース制御手段は、受け取ったプリフェッチ・アドレスに対応したキャッシュブロックのフェッチを行うとともに、前記主プロセッサが次に行うと予想されるアクセスに備えて待機することを特徴とするソフトウェア制御可能なキャッシュ・メモリ装置の制御方法。

【請求項12】 前記プリフェッチを行うために、前記主プロセッサが前記キャッシュ・メモリに対してメモリ・アクセス要求を行う処理ステップと、前記アクセス要求をアドレス・バスとプリフェッチ・コード・セレクト信号を通じて前記専用プロセッサに伝える処理ステップと、前記専用プロセッサが、前記主プロセッサからプリフェッチ・コード・セレクト信号を受け取ると、前記プリフ

エッチ・コード・エントリ・アドレス・テーブルを用いて前記プリフェッチ・コード・メモリにある命令列の先頭アドレスを検索する処理ステップと、

前記検索を受けた前記プリフェッチ・コード・エントリ・アドレス・テーブルが、前記プリフェッチ・コード・セレクト信号に応答して、対応する命令列の先頭アドレスを返す処理ステップと、

前記専用プロセッサが前記検索で得られた命令から命令の実行を開始する処理ステップと、

前記専用プロセッサが、前記プリフェッチ・アドレスを生成する命令を実行した場合は、前記フェッチ／リプレース制御手段に対してキャッシュ・アクセス・アドレス・バスを通じてプリフェッチすべきアドレスを送出する処理ステップと、

前記専用プロセッサが、受け取ったアドレスに対応するキャッシュ・ブロックを入れ替える処理ステップと、

前記プリフェッチ・コードの実行が終了するか、あるいは前記主プロセッサが前記プリフェッチ・コード・セレクト信号を通じてプリフェッチの停止指示を受けてプリフェッチを停止する処理ステップと、を有する請求項1記載のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法。

【請求項13】 プリフェッチを行わない場合は、前記主プロセッサが、前記キャッシュ・メモリに対してメモリアccessを要求し、かつ、前記専用プロセッサに対してプリフェッチを行わないように、前記プリフェッチ・コード・セレクト信号を通じて指示する処理ステップと、

指示を受けた前記専用プロセッサが、プリフェッチを行わず単に前記キャッシュ・メモリアccessを行う処理ステップとを有する請求項1記載のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法。

【請求項14】 前記プリフェッチ・コードの読み込みおよび入れ替えを行うために、前記主プロセッサが、前記専用プロセッサに対してプリフェッチ・コード・メモリアccessを、前記プリフェッチ・コード・セレクト信号を通じて指示する処理ステップと、

前記専用プロセッサが、受け取ったアドレスに基づき前記主メモリからプリフェッチ・コードを読み出す処理ステップと、

前記主プロセッサが、読み出すサイズと読み出したプリフェッチ・コードに対応するプリフェッチ指示コードを、前記専用プロセッサに対してデータ・バスを通じて与える処理ステップと、

前記専用プロセッサが、前記主メモリから読み出した前記プリフェッチ・コードを前記プリフェッチ・コード・メモリにロードする処理ステップと、

前記ロードと同時に、ロードした命令列の先頭アドレスと前記主プロセッサから与えられたプリフェッチ指示コードとを前記プリフェッチ・コード・エントリ・アドレ

ス・テーブルに登録する処理ステップと、を有する請求項1記載のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法。

【請求項15】 コンパイラのコード生成のために、前記コンパイラが、ループまたはリンク・ポインタの追跡処理のようなプリフェッチが有効であるコードを認識すると、前記主メモリアccess・パターンを解析し、かつ前記解析結果に基づき前記主メモリアccess・パターンについて有効なプリフェッチ・パターンを推測する処理ステップと、

前記主メモリアccess・パターンを解析し、前記主プロセッサの暫定的な命令列を生成する処理ステップと、推測した前記プリフェッチ・パターンを、前記専用プロセッサが出力する前記プリフェッチ・コードとして生成する処理ステップと、

前記有効なプリフェッチ・パターンが複数推測される場合は、それぞれについて前記プリフェッチ・コードを生成する処理ステップと、

生成した前記プリフェッチ・コードに対して、生成プログラム中で一意となるプリフェッチ指示コードを定義する処理ステップと、

前記コンパイラが、前記有効なプリフェッチ・パターンを推測する処理ステップで生成した前記主プロセッサ用の命令のうちプリフェッチ・コードを利用するメモリアccess命令を、メモリアccessとともにプリフェッチも行う命令に置き換える処理ステップと、

置き換える際に、前記専用プロセッサが出力する前記プリフェッチ・コードとして生成する処理ステップで生成したプリフェッチ・コードのうち、もっともプリフェッチの効果が高いと予測されるプリフェッチ・コードに対応したプリフェッチ指示コードを前記専用プロセッサに与えるようにメモリアccess命令を置き換える処理ステップと、を有する請求項1記載のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法。

【請求項16】 前記プリフェッチ・パターンの解析は、ユーザがソースコード中に埋め込むコンパイラ・ディレクティブによって指定する請求項15記載のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法。

【請求項17】 前記プリフェッチ・コードメモリに格納されるコードのみを入れ替えることで、プリフェッチ・パターンを変更する請求項15記載のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法。

【請求項18】 プリフェッチのパターンを複数設定し、かつそれらを選択して所定のプログラムに対してプリフェッチを行って、前記主メモリがデータを要求してから使用できるまでの平均的なメモリアccessのレイテンシを隠蔽する請求項15記載のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法。

【請求項19】 プログラムを再コンパイルすることなくプログラムのチューニングを行う請求項15記載のソ

フトウェア制御可能なキャッシュ・メモリ装置の制御方法。

【請求項20】 主プロセッサのキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ・リプレース制御手段とを用いて、前記プリフェッチ・アドレスの決定をプログラム実行時に行うために、前記専用プロセッサが、前記プリフェッチ・コード・メモリに格納されている命令を読み出し逐次実行してプリフェッチ・アドレスを生成し、生成したプリフェッチ・アドレスを前記フェッチ／リプレース制御手段に対して送出し、前記フェッチ／リプレース制御手段は、受け取ったプリフェッチ・アドレスに対応したキャッシュブロックのフェッチを行うとともに、前記主プロセッサが次に行うと予想されるアクセスに備えて待機する機能を、前記主プロセッサに実行させるプログラム。

【請求項21】 主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ・リプレース制御手段とを用いて、前記プリフェッチを行うために、前記主プロセッサが前記キャッシュ・メモリに対してメモリ・アクセス要求を行い、前記アクセス要求をアドレス・バスとプリフェッチ・コード・セレクト信号を通じて前記専用プロセッサに伝えると、

前記専用プロセッサが、前記主プロセッサからプリフェッチ・コード・セレクト信号を受け取り、前記プリフェッチ・コード・エントリ・アドレス・テーブルを用いて前記プリフェッチ・コード・メモリにある命令列の先頭アドレスを検索する処理ステップと、

前記検索を受けた前記プリフェッチ・コード・エントリ・アドレス・テーブルが、前記プリフェッチ・コード・セレクト信号に応答して、対応する命令列の先頭アドレスを返す処理ステップと、

前記専用プロセッサが前記検索で得られた命令から命令の実行を開始する処理ステップと、

前記専用プロセッサが、前記プリフェッチ・アドレスを生成する命令を実行した場合は、前記フェッチ／リプレ

ース制御手段に対してキャッシュ・アクセス・アドレス・バスを通じてプリフェッチすべきアドレスを送出する処理ステップと、

前記専用プロセッサが、受け取ったアドレスに対応するキャッシュ・ブロックを入れ替える処理ステップと、

前記プリフェッチ・コードの実行が終了するか、あるいは前記主プロセッサが前記プリフェッチ・コード・セレクト信号を通じてプリフェッチの停止指示を受けてプリフェッチを停止する処理ステップと、を前記専用プロセッサに実行させるためのプログラム。

【請求項22】 主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ・リプレース制御手段とを用いて、

プリフェッチを行わない場合は、前記主プロセッサが、前記キャッシュ・メモリに対してメモリアccessを要求し、かつ、前記専用プロセッサに対してプリフェッチを行わないように、前記プリフェッチ・コード・セレクト信号を通じて指示することにより、指示を受けた前記専用プロセッサがプリフェッチを行わず単に前記キャッシュ・メモリのアクセスを行うようにする処理ステップを、前記主プロセッサに実行させるためのプログラム。

【請求項23】 主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ・リプレース制御手段とを用いて、前記プリフェッチ・コードの読み込みおよび入れ替えを行うために、前記主プロセッサが、前記専用プロセッサに対してプリフェッチ・コード・メモリの入れ替えを、前記プリフェッチ・コード・セレクト信号を通じて指示すると、

前記専用プロセッサが、受け取ったアドレスに基づき前記主メモリからプリフェッチ・コードを読み出す処理ステップと、

前記主プロセッサが、読み出すサイズと読み出したプリフェッチ・コードに対応するプリフェッチ指示コードを、前記専用プロセッサに対してデータ・バスを通じて与えると、前記専用プロセッサが、前記主メモリから読

み出した前記ブリフェッチ・コードを前記ブリフェッチ・コード・メモリにロードする処理ステップと、前記ロードと同時に、ロードした命令列の先頭アドレスと前記主プロセッサから与えられたブリフェッチ指示コードとを前記ブリフェッチ・コード・エントリ・アドレス・テーブルに登録する処理ステップと、を前記専用プロセッサに実行させるためのプログラム。

【請求項24】 主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するブリフェッチ・コード・メモリと、前記主プロセッサからのブリフェッチ指示内容と前記ブリフェッチ・コード・メモリのエントリアドレスとを対で格納するブリフェッチ・コード・エントリ・アドレス・テーブルとを有するブリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ・リプレース制御手段とを用いて、コンパイラのコード生成のために、前記コンパイラが、ループまたはリンク・ポインタの追跡処理のようなブリフェッチが有効であるコードを認識すると、前記主メモリのアクセス・パターンを解析し、かつ前記解析結果に基づき前記主メモリのアクセス・パターンについて有効なブリフェッチ・パターンを推測する処理ステップと、前記主メモリのアクセス・パターンを解析し、前記主プロセッサの暫定的な命令列を生成する処理ステップと、推測した前記ブリフェッチ・パターンを、前記専用プロセッサが出力する前記

ブリフェッチ・コードとして生成する処理ステップと、前記有効なブリフェッチ・パターンが複数推測される場合は、それぞれについて前記ブリフェッチ・コードを生成する処理ステップと、

生成した前記ブリフェッチ・コードに対して、生成プログラム中で一意となるブリフェッチ指示コードを定義する処理ステップと、

前記コンパイラが、前記有効なブリフェッチ・パターンを推測する処理ステップで生成した前記主プロセッサ用の命令のうちブリフェッチ・コードを利用するメモリアクセス命令を、メモリアクセスとともにブリフェッチも行う命令に置き換える処理ステップと、

置き換える際に、前記専用プロセッサが出力する前記ブリフェッチ・コードとして生成する処理ステップで生成したブリフェッチ・コードのうち、もっともブリフェッチの効果が高いと予測されるブリフェッチ・コードに対応したブリフェッチ指示コードを前記専用プロセッサに与えるようにメモリアクセス命令を置き換える処理ステップと、をコンパイラシステムに実行させるためのプログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明はソフトウェア制御可

能なキャッシュ・メモリ装置、制御方法およびプログラムに係わり、特にキャッシュ・メモリを有するプロセッサがブリフェッチ・アドレスの生成をソフトウェアによって制御するブリフェッチ・アドレス生成用プロセッサを備えることにより、ソフトウェアで制御可能なブリフェッチ機構を実現するソフトウェア制御可能なキャッシュ・メモリ装置、制御方法およびプログラムに関する。

【0002】

【従来の技術】 近年、半導体素子の微細化技術の進展に伴い、その半導体素子で構成するLSIも大規模化、高速化している。特に中央演算処理装置としてのマイクロプロセッサの高速化とともに、対応する記憶装置の大規模化も著しく、例えば、外部記憶装置に用いる半導体記憶装置では1チップに256メガバイトの容量を有するダイナミック型ランダムアクセスメモリも実用化されている。また、これらの微細化技術はマイクロコンピュータおよびマイクロコンピュータに搭載される内部記憶装置にも適用され、記憶容量の増大化に貢献している。

【0003】 一方、上述したマイクロプロセッサやマイクロコンピュータを搭載する応用システムにおいて、キャッシュ容量も上述した半導体素子の微細化技術の進展によって、キロビットからメガビット単位に増大している。増大したキャッシュを意識したプログラムの最適化技術も性能が向上してきているが、実行前のプログラムからメモリアクセスを予測してキャッシュヒット率を向上させる試みも種々行われている。

【0004】 キャッシュは、マイクロプロセッサと同一半導体基板上に構成される場合もあるが、その周辺に配置されることもある主メモリ等比べて小容量に構成された高速のメモリである。マイクロプロセッサが最も近い時期に使用したデータおよび命令はキャッシュ内に保存されているが、マイクロプロセッサにより再度アクセスされない限り、ある時間が経過すると主メモリに再び書き戻される。

【0005】 キャッシュがマイクロプロセッサによってアクセスされるとき、アドレスの一部がインデックスされるが、インデックスされたメモリラインに主メモリの所望のアドレスに記憶されたデータまたは命令が含まれているときに、アクセスがヒットしたと称する。逆に、インデックスされたメモリラインに主メモリの所望のアドレスに記憶されたデータまたは命令が含まれていないときは、キャッシュミスしたと称する。

【0006】 一方、コンパイラの解析結果を利用しソフトウェアでキャッシュを制御する方法も提案されている。このソフトウェアによるキャッシュ制御の一例として、将来使用するであろうと予測されるデータを予めキャッシュに記憶しておく方法がある。

【0007】 例えば、ブリフェッチ機構を有する方式の一例を示した図9を参照すると、ブリフェッチするアドレスをハードウェアによって生成する方式（以下、方式

Aと称す)である。実現方法には様々な方式があるが、ハードウェアの実装上の制約から、基底アドレスからストライドが一定となるアドレス列をプリフェッチするものが主流であり、プリフェッチ・アドレスは、以下のような数列で表される。

$$P_n = B + S \times n$$

$P_n$ : プリフェッチ・アドレス

$B$ : 基底アドレス

$S$ : ストライド

$n$ : プリフェッチされる順序番号

これを行うために、通常のキャッシュに加えて以下のハードウェア機構が必要となる。

【0008】従来の上述した方式Aのキャッシュ・メモリ装置は、主プロセッサ700と主メモリ900との間に、キャッシュ・メモリ800が設けられる。

【0009】キャッシュ・メモリ800は、プリフェッチしたアドレスを記録するフェッチ・アドレス・レジスタ801と、ストライドを記録するアドレス差分レジスタ802と、次のプリフェッチするアドレスを生成する加減算器803と、フェッチアドレスの選択を行うアドレス・セクタ804および805と、プリフェッチおよびプロセッサからの通常アクセスの調停を行うフェッチ制御回路806と、フェッチ/リブレース制御部807と、キャッシュ・アレイ808とを備える。

【0010】上述した構成からなるプリフェッチ機構を備えたキャッシュ・メモリ装置におけるプリフェッチは以下の手順で行われる。

【0011】まず、主プロセッサ700あるいはフェッチ制御回路806が、アドレス差分レジスタ802にストライドを設定する(処理1)。

【0012】次に、主プロセッサ700あるいはフェッチ制御回路806が、プリフェッチの開始アドレスとなる基底アドレスをフェッチ・アドレス・レジスタ801に設定する(処理2)。

【0013】フェッチ・アドレスが増加数列を成すか減少数列を成すかを判断する(処理3)。

【0014】フェッチ・アドレスが増加数列を成すならば、加算方向指定信号815を通じて加減算器803が加算処理を行うようにする(処理4)。

【0015】逆にフェッチ・アドレスが減少数列を成すならば加算方向指定信号815を通じて加減算器803が減算処理を行うようにする(処理5)。

【0016】主プロセッサ700あるいはフェッチ制御回路806がプリフェッチ処理を開始する(処理6)。これによりセクタ805は、フェッチ・アドレス・レジスタ801の入力としてプリフェッチ・アドレス817を選択する(処理7)。

【0017】同時に、フェッチ/リブレース制御部807への入力として加減算器803からの出力を選択するようにアドレス・セクタ804を設定する(処理

8)。

【0018】これによりフェッチ/リブレース制御部807にプリフェッチすべきアドレスを入力する(処理9)。

【0019】フェッチ/リブレース制御部807は与えられたアドレスに基づいてキャッシュのタグ・アレイ808、フラグ・アレイ809、データ・アレイ810を更新する(処理10)。

【0020】主プロセッサ700あるいはフェッチ制御回路806の指示により、プリフェッチする必要がなくなるまで処理3～処理10の処理ステップを繰り返す。

【0021】主プロセッサ700からの通常のアクセスを行う場合は、プリフェッチ指示信号711を通じてフェッチ制御回路806によるプリフェッチ処理の抑止を行い、プリフェッチ処理をバイパスさせて直接フェッチ/リブレース制御部807にアドレスを渡すことで実現できる。

【0022】なお、プリフェッチ開始アドレスやストライドの設定および検出方法については各種方式が提案されているが、それらは本提案には本質的に異なる方式であるので、それらについての説明は省略する。

【0023】一方、他の従来例として、キャッシュ・メモリを有するプロセッサコアにおいて、プロセッサコアがデータを要求してから実際に使用するまでの平均的なレイテンシを小さく押さえる例が特開2000-172562号公報に記載されている。同公報記載の情報処理装置では、プリフェッチ/ポストストアのアドレスをソフトウェアから指示するための特殊レジスタとこの特殊レジスタによって指示されたパラメータから物理アドレスを生成するアドレス生成変換装置と、全体の制御を行うプリフェッチ/ポストストアコントローラを含むプリフェッチ/ポストストア回路を備え、プリフェッチに必要な情報をハードウェアおよびソフトウェアの協働によって実現し、プリフェッチ情報は、プロセッサ上で動作するソフトウェアの指示により、プリフェッチ/ポストストア回路の特殊レジスタにソフトウェアによって設定される。

【0024】このプリフェッチ/ポストストアは、プロセッサによるキャッシュ・メモリ制御を必要とせず、かつプロセッサが管理する情報を書き換えることなく実行されるため、プロセッサにおける処理やプロセッサとメモリアクセスコントローラとの間の低速なインタフェースによって律速されない、高速処理が可能というものである。

【0025】

【発明が解決しようとする課題】上述したように従来のプリフェッチ機構を備えたソフトウェア制御可能なキャッシュ・メモリ装置は、データベース処理等のアプリケーションを適用した場合は、プリフェッチが有効に働かないことが多い。

【0026】特に配列に対する処理が多い数値演算アプリケーションでは、ハードウェアによるプリフェッチは有効であるが、データベース処理などの事務系のアプリケーションでは、データの格納方式としてリスト構造やツリー構造を用いることが多く、プリフェッチを適用できない。

【0027】リスト構造やツリー構造では、データ要素間の関連をポインタとして表現するが、これらデータ構造は以下のような性質を持つことが知られている。

1. 関連するデータ要素群が連続したアドレス領域に配置されることが多くない。
2. データ要素の処理が順次行われることが多い。このため一度参照されたデータが狭いタイムスケールの範囲で再びアクセスされる可能性が低い。

【0028】上述した1項は、すなわち、アクセスパターンに空間局所性がないことを意味し、2項は、すなわち、アクセスパターンに時間局所性がないことを意味する。

【0029】このため、事務系アプリケーションでは、キャッシュのミスヒット率が増加することとなり、主メモリのアクセスレイテンシがプロセッサの実効性能を低減させることとなる。

【0030】方式Aのプリフェッチ機構を備えたソフトウェア制御可能なキャッシュ・メモリ装置がプリフェッチするには、プリフェッチするアドレス列が一定のストライドであることを要求する。

【0031】しかしながら、リスト構造やツリー構造を持つデータでは、データのアクセス順序にしたがって各要素を一定間隔で配置することは原理的に難しい。すなわちリスト構造やツリー構造のデータでは、方式Aの求めるプリフェッチ条件を満たせないのである。

【0032】これを改善する方法として、プロセッサが明示的にプリフェッチ命令を発行し、キャッシュがプリフェッチ命令に従ってプリフェッチを行う方式（以下、方式Bと称す）もある。

【0033】この方式Bでは、プロセッサがプリフェッチするアドレスを指定するため、プリフェッチ・アドレス列は一定のストライドでなくてもよい。

【0034】しかし、方式Bも問題がある。通常、プリフェッチすべきアドレスの決定処理と、実際にデータのアクセスを行う処理とは、時系列上で非常に接近していることが多い。

【0035】プリフェッチするアドレスをプロセッサが決定する以上、プリフェッチ命令と実際のデータアクセス処理もまた時系列上で接近することとなる。

【0036】このため、プリフェッチのレイテンシが大きい場合には方式Bであっても、キャッシュミスのレイテンシを隠蔽することは不可能であることが多い。

【0037】また、方式Bにおけるプリフェッチ命令は、プログラムの状態を更新しない命令であり、プログ

ラムの処理には何ら関係がない。

【0038】プロセッサの性能をデータ処理速度という観点で捉えるならば、プリフェッチ命令は無駄な命令である。

【0039】プロセッサがハーバードアーキテクチャを取るのであれば、プリフェッチ命令によってデータキャッシュのヒット率が上がるが、命令キャッシュ内の有効データ密度を低減させる、というトレードオフが存在する。

【0040】このように、既存のプリフェッチ実現方式ではリスト構造やツリー構造などに代表されるランダム性を伴うアクセスパターンでは、プリフェッチの効果が得にくいという問題がある。

【0041】一方、特開2000-172562号公報の従来例は、後述する本発明の特徴である、アドレス生成に特化したプリフェッチ・アドレス生成用プロセッサと、このプリフェッチ・アドレス生成用プロセッサにプリフェッチを行わせる命令を格納するプリフェッチ・コード・メモリと、プリフェッチ・アドレス生成用プロセッサからのプリフェッチ指示内容とプリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ／リプレース制御手段とを有する構成とは異なり、かつプリフェッチ・コード・メモリに格納されるコードのみを入れ替えることで、主プロセッサに対するコードを変更することなくプリフェッチ・パターンを変更する作用効果も実現することが出来ない。

【0042】本発明の目的は、上述した従来の欠点に鑑みなされたものであり、キャッシュ・メモリを有するプロセッサがプリフェッチ・アドレスの生成をソフトウェアによって制御するプリフェッチ・アドレス生成用プロセッサを備えることにより、ソフトウェアで制御可能なプリフェッチ機構を実現することにある。

【0043】

【課題を解決するための手段】本発明のソフトウェア制御可能なキャッシュ・メモリ装置は、主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリが、プログラム実行時に、プリフェッチ・アドレスをソフトウェア制御で決定するプリフェッチ手段として、実行すべきプリフェッチ・コードの指定を前記主プロセッサから指示されるとともに前記プリフェッチ・コードに対応した命令の実行により前記プリフェッチ・アドレスを生成する専用プロセッサを含み、かつ前記主プロセッサと並列処理を行うプリフェッチ・エンジンを備えたことを特徴とする。

【0044】また、前記専用プロセッサは、浮動小数点演算機能および割り込み機能が省略された特定用途のプロセッサである。

【0045】さらに、前記専用プロセッサは、前記主プロセッサからの指示に従い、プリフェッチ・コード・メモリに格納されている専用プロセッサ自身の制御用命令を読み出し、読み出した前記命令を自律的に逐次実行して前記プリフェッチ・アドレスを生成するとともに、生成した前記プリフェッチ・アドレスを前記キャッシュ・メモリ内にあるフェッチ／リブレース制御手段に対して送出する機能を有する。

【0046】本発明のソフトウェア制御可能なキャッシュ・メモリ装置の他の特徴は、主プロセッサのキャッシュ・メモリ内に、アドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記専用プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ／リブレース制御手段と備えることにある。

【0047】また、前記プリフェッチ・コード・メモリを設ける代わりに、主メモリにプリフェッチ・コードを格納し、主メモリを主プロセッサと専用プロセッサで共有することもできる。

【0048】さらに、前記プリフェッチ・コード・メモリを、前記プリフェッチコードメモリに変えて、前記キャッシュメモリ内に格納することもできる。

【0049】さらにまた、前記フェッチ／リブレース制御手段は、前記専用プロセッサから受け取ったプリフェッチ・アドレスに対応するキャッシュ・ブロックをフェッチして入れ替えるとともに、前記主プロセッサが次に実行するアクセスまで待機する機能を有してもよい。

【0050】また、前記プリフェッチ・エンジン作業用レジスタ・ファイルは、前記専用プロセッサが必要とするデータ全てを格納する機能を有し、前記専用プロセッサは前記主メモリを直接アクセスせずに前記プリフェッチ・エンジン作業用レジスタ・ファイルをアクセスする機能を有する。

【0051】さらに、前記主プロセッサは、前記プリフェッチ・コード・メモリに格納された、前記専用プロセッサが実行すべき命令列を指定するためのプリフェッチ指示コードをもつプリフェッチ・コード・セレクト信号を有し、そのプリフェッチ・コード・セレクト信号を前記専用プロセッサに転送する。

【0052】さらにまた、前記プリフェッチ・コード・セレクト信号のビット幅を、メモリアクセス命令にプリフェッチ・コードのアドレスを指定できるだけの情報に必要なビット幅に予め拡張して備える。

【0053】本発明のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法は、主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内にアドレス生

成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ・リブレース制御手段と備え、前記プリフェッチ・アドレスの決定をプログラム実行時に行うために、前記専用プロセッサが、前記プリフェッチ・コード・メモリに格納されている命令を読み出し逐次実行してプリフェッチ・アドレスを生成し、生成したプリフェッチ・アドレスを前記フェッチ／リブレース制御手段に対して送出し、前記フェッチ／リブレース制御手段は、受け取ったプリフェッチ・アドレスに対応したキャッシュブロックのフェッチを行うとともに、前記主プロセッサが次に行うと予想されるアクセスに備えて待機することを特徴とする。

【0054】また、前記プリフェッチを行うために、前記主プロセッサが前記キャッシュ・メモリに対してメモリ・アクセス要求を行う処理ステップと、前記アクセス要求をアドレス・バスとプリフェッチ・コード・セレクト信号を通じて前記専用プロセッサに伝える処理ステップと、前記専用プロセッサが、前記主プロセッサからプリフェッチ・コード・セレクト信号を受け取ると、前記プリフェッチ・コード・エントリ・アドレス・テーブルを用いて前記プリフェッチ・コード・メモリにある命令列の先頭アドレスを検索する処理ステップと、前記検索を受けた前記プリフェッチ・コード・エントリ・アドレス・テーブルが、前記プリフェッチ・コード・セレクト信号に応答して、対応する命令列の先頭アドレスを返す処理ステップと、前記専用プロセッサが前記検索で得られた命令から命令の実行を開始する処理ステップと、前記専用プロセッサが、前記プリフェッチ・アドレスを生成する命令を実行した場合は、前記フェッチ／リブレース制御手段に対してキャッシュ・アクセス・アドレス・バスを通じてプリフェッチすべきアドレスを送出する処理ステップと、前記専用プロセッサが、受け取ったアドレスに対応するキャッシュ・ブロックを入れ替える処理ステップと、前記プリフェッチ・コードの実行が終了するか、あるいは前記主プロセッサが前記プリフェッチ・コード・セレクト信号を通じてプリフェッチの停止指示を受けてプリフェッチを停止する処理ステップと、を有する。

【0055】さらに、プリフェッチを行わない場合は、前記主プロセッサが、前記キャッシュ・メモリに対してメモリアクセスを要求し、かつ、前記専用プロセッサに対してプリフェッチを行わないように、前記プリフェッチ・コード・セレクト信号を通じて指示する処理ステップと、指示を受けた前記専用プロセッサが、プリフェ



チを行わず単に前記キャッシュ・メモリのアクセスを行う処理ステップとを有する。

【0056】さらにまた、前記プリフェッチ・コードの読み込みおよび入れ替えを行うために、前記主プロセッサが、前記専用プロセッサに対してプリフェッチ・コード・メモリの入れ替えを、前記プリフェッチ・コード・セレクト信号を通じて指示する処理ステップと、前記専用プロセッサが、受け取ったアドレスに基づき前記主メモリからプリフェッチ・コードを読み出す処理ステップと、前記主プロセッサが、読み出すサイズと読み出したプリフェッチ・コードに対応するプリフェッチ指示コードを、前記専用プロセッサに対してデータ・バスを通じて与える処理ステップと、前記専用プロセッサが、前記主メモリから読み出した前記プリフェッチ・コードを前記プリフェッチ・コード・メモリにロードする処理ステップと、前記ロードと同時に、ロードした命令列の先頭アドレスと前記主プロセッサから与えられたプリフェッチ指示コードとを前記プリフェッチ・コード・エントリ・アドレス・テーブルに登録する処理ステップと、を有する。

【0057】また、コンパイラのコード生成のために、前記コンパイラが、ループまたはリンク・ポインタの追跡処理のようなプリフェッチが有効であるコードを認識すると、前記主メモリのアクセス・パターンを解析し、かつ前記解析結果に基づき前記主メモリのアクセス・パターンについて有効なプリフェッチ・パターンを推測する処理ステップと、前記主メモリのアクセス・パターンを解析し、前記主プロセッサの暫定的な命令列を生成する処理ステップと、推測した前記プリフェッチ・パターンを、前記専用プロセッサが出力する前記プリフェッチ・コードとして生成する処理ステップと、前記有効なプリフェッチ・パターンが複数推測される場合は、それぞれについて前記プリフェッチ・コードを生成する処理ステップと、生成した前記プリフェッチ・コードに対して、生成プログラム中で一意となるプリフェッチ指示コードを定義する処理ステップと、前記コンパイラが、前記有効なプリフェッチ・パターンを推測する処理ステップで生成した前記主プロセッサ用の命令のうちプリフェッチ・コードを利用するメモリアccess命令を、メモリアccessとともにプリフェッチも行う命令に置き換える処理ステップと、置き換える際に、前記専用プロセッサが出力する前記プリフェッチ・コードとして生成する処理ステップで生成したプリフェッチ・コードのうち、もっともプリフェッチの効果が高いと予測されるプリフェッチ・コードに対応したプリフェッチ指示コードを前記専用プロセッサに与えるようにメモリアccess命令を置き換える処理ステップと、を有する。

【0058】また、前記プリフェッチ・パターンの解析は、ユーザがソースコード中に埋め込むコンパイラ・ディレクティブによって指定することができる。

【0059】さらに、前記プリフェッチ・コードメモリに格納されるコードのみを入れ替えることで、プリフェッチ・パターンを変更することもできる。

【0060】さらにまた、プリフェッチのパターンを複数設定し、かつそれらを選択して所定のプログラムに対してプリフェッチを行って、前記主メモリがデータを要求してから使用できるまでの平均的なメモリアccessのレイテンシを隠蔽することもできる。

【0061】プログラムを再コンパイルすることなくプログラムのチューニングを行うこともできる。

【0062】本発明のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法のプログラムは、主プロセッサのキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ・リプレース制御手段と備え、前記プリフェッチ・アドレスの決定をプログラム実行時に行うために、前記専用プロセッサが、前記プリフェッチ・コード・メモリに格納されている命令を読み出し逐次実行してプリフェッチ・アドレスを生成し、生成したプリフェッチ・アドレスを前記フェッチ／リプレース制御手段に対して送出し、前記フェッチ／リプレース制御手段は、受け取ったプリフェッチ・アドレスに対応したキャッシュブロックのフェッチを行うとともに、前記主プロセッサが次に行うと予想されるアクセスに備えて待機する機能を、前記主プロセッサに実行させるプログラムを備えることを特徴とする。

【0063】本発明のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法のプログラム他の特徴は、主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ・リプレース制御手段とを用いて、前記プリフェッチを行うために、前記主プロセッサが前記キャッシュ・メモリに対してメモリ・アクセス要求を行い、前記アクセス要求をアドレス・バスとプリフェッチ・コード・セレクト信号を通じて前記専用プロセッサに伝えると、前記専用プロセッサが、前記主プロセッサからプリフェッチ・コード・セレクト信号を受け取ると、前記プリフェッチ・コー

ド・エントリ・アドレス・テーブルを用いて前記プリフェッチ・コード・メモリにある命令列の先頭アドレスを検索する処理ステップと、前記検索を受けた前記プリフェッチ・コード・エントリ・アドレス・テーブルが、前記プリフェッチ・コード・セレクト信号に応答して、対応する命令列の先頭アドレスを返す処理ステップと、前記専用プロセッサが前記検索で得られた命令から命令の実行を開始する処理ステップと、前記専用プロセッサが、前記プリフェッチ・アドレスを生成する命令を実行した場合は、前記フェッチ／リブレース制御手段に対してキャッシュ・アクセス・アドレス・バスを通じてプリフェッチすべきアドレスを送出する処理ステップと、前記専用プロセッサが、受け取ったアドレスに対応するキャッシュ・ブロックを入れ替える処理ステップと、前記プリフェッチ・コードの実行が終了するか、あるいは前記主プロセッサが前記プリフェッチ・コード・セレクト信号を通じてプリフェッチの停止指示を受けてプリフェッチを停止する処理ステップと、を前記専用プロセッサに実行させるためのプログラムを備えることにある。

【0064】本発明のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法のプログラムまた他の特徴は、主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ・リブレース制御手段とを用いて、プリフェッチを行わない場合は、前記主プロセッサが、前記キャッシュ・メモリに対してメモリアccessを要求し、かつ、前記専用プロセッサに対してプリフェッチを行わないように、前記プリフェッチ・コード・セレクト信号を通じて指示することにより、指示を受けた前記専用プロセッサがプリフェッチを行わず単に前記キャッシュ・メモリのアクセスを行うようにする処理ステップを、前記主プロセッサに実行させるためのプログラムを備えることにある。

【0065】本発明のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法のプログラムのさらにまた他の特徴は、主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフ

ェッチ・リブレース制御手段とを用いて、前記プリフェッチ・コードの読み込みおよび入れ替えを行うために、前記主プロセッサが、前記専用プロセッサに対してプリフェッチ・コード・メモリの入れ替えを、前記プリフェッチ・コード・セレクト信号を通じて指示すると、前記専用プロセッサが、受け取ったアドレスに基づき前記主メモリからプリフェッチ・コードを読み出す処理ステップと、前記主プロセッサが、読み出すサイズと読み出したプリフェッチ・コードに対応するプリフェッチ指示コードを、前記専用プロセッサに対してデータ・バスを通じて与えると、前記専用プロセッサが、前記主メモリから読み出した前記プリフェッチ・コードを前記プリフェッチ・コード・メモリにロードする処理ステップと、前記ロードと同時に、ロードした命令列の先頭アドレスと前記主プロセッサから与えられたプリフェッチ指示コードとを前記プリフェッチ・コード・エントリ・アドレス・テーブルに登録する処理ステップと、を前記専用プロセッサに実行させるためのプログラムを備えることにある。

【0066】本発明のソフトウェア制御可能なキャッシュ・メモリ装置の制御方法のプログラムの他の特徴は、主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリ内にアドレス生成に特化した専用プロセッサと、前記専用プロセッサの命令を格納するプリフェッチ・コード・メモリと、前記主プロセッサからのプリフェッチ指示内容と前記プリフェッチ・コード・メモリのエントリアドレスとを対で格納するプリフェッチ・コード・エントリ・アドレス・テーブルとを有するプリフェッチ・エンジンと、前記専用プロセッサからの指示アドレスに従いキャッシュ・ブロックを入れ替えるフェッチ・リブレース制御手段とを用いて、コンパイラのコード生成のために、前記コンパイラが、ループまたはリンク・ポイントの追跡処理のようなプリフェッチが有効であるコードを認識すると、前記主メモリのアクセス・パターンを解析し、かつ前記解析結果に基づき前記主メモリのアクセス・パターンについて有効なプリフェッチ・パターンを推測する処理ステップと、前記主メモリのアクセス・パターンを解析し、前記主プロセッサの暫定的な命令列を生成する処理ステップと、推測した前記プリフェッチ・パターンを、前記専用プロセッサが出力する前記プリフェッチ・コードとして生成する処理ステップと、前記有効なプリフェッチ・パターンが複数推測される場合は、それぞれについて前記プリフェッチ・コードを生成する処理ステップと、生成した前記プリフェッチ・コードに対して、生成プログラム中で一意となるプリフェッチ指示コードを定義する処理ステップと、前記コンパイラが、前記有効なプリフェッチ・パターンを推測する処理ステップで生成した前記主プロセッサ用の命令のうちプリフェッチ・コードを利用するメモリアccess命令を、メモリアccessとともにプリフェッチも行う命令に

置き換える処理ステップと、置き換える際に、前記専用プロセッサが出力する前記ブリフェッチ・コードとして生成する処理ステップで生成したブリフェッチ・コードのうち、もっともブリフェッチの効果が高いと予測されるブリフェッチ・コードに対応したブリフェッチ指示コードを前記専用プロセッサに与えるようにメモリアクセス命令を置き換える処理ステップと、を前記主プロセッサに実行させるためのプログラムを備えることにある。

【0067】

【発明の実施の形態】本発明のソフトウェア制御可能なキャッシュ・メモリ装置は、キャッシュ内にブリフェッチ・エンジンを設けることで、フェッチアドレスの生成および発行に柔軟性をもたせるものである。

【0068】まず、本発明の第1の実施形態を図面を参照しながら詳細に説明する。

【0069】第1の実施形態の主要部の構成を示した図1を参照すると、ソフトウェア制御可能なキャッシュ・メモリ装置200（以下、キャッシュ・メモリ装置と称す）は、主プロセッサ100と主メモリ300との間に設けられる。

【0070】キャッシュ・メモリ装置200は、キャッシュ・アレイ400とブリフェッチ・エンジン500とブリフェッチ／リブレース制御部600とを備える。

【0071】キャッシュ・アレイ400は、タグ・アレイ401とフラグ・アレイ402とデータ・アレイ403とを備え、ブリフェッチ・アドレス生成用プロセッサ501との間でブリフェッチすべきアドレスの受け渡しに用いられるブリフェッチ・アドレス・バス122で接続され、ブリフェッチ／リブレース制御部600との間では、タグ・アレイ401との間を双方向のタグ・データ・バス211で接続され、フラグ・アレイ402との間では双方向のフラグ・データ・バス212で接続され、データ・アレイ403との間では双方向のブロック・データ・バス213で接続される。

【0072】ブリフェッチ・エンジン500は、アドレス生成処理に特化した高速小型のプロセッサであるブリフェッチ・アドレス生成用プロセッサ（請求項における専用プロセッサに対応）501と、このブリフェッチ・アドレス生成用プロセッサ501の命令を格納するためのメモリであるブリフェッチ・コード・メモリ504とを備え、両者間は双方向のブリフェッチ・コード・データ・バス511で接続され、さらにブリフェッチ・アドレス生成用プロセッサ501からの単一方向のブリフェッチ・コード・アドレス・バス512で接続される。

【0073】ブリフェッチ・アドレス生成用プロセッサ501は、一般的なプロセッサから、浮動小数点演算機能が省略された特定用途のプロセッサである。

【0074】一方、フェッチ／リブレース制御部600との間では双方向のキャッシュ・アクセス・アドレス・バス214およびブリフェッチ調停信号215で接続さ

れる。

【0075】ブリフェッチ・アドレス生成用プロセッサ501は、ブリフェッチ・エンジン作業用レジスタ・ファイル502と、主プロセッサ100から与えられるブリフェッチ指示内容とブリフェッチ・コード・メモリ504のエントリ・アドレスを格納するブリフェッチ・コード・エントリ・アドレス・テーブル503とからなり、このブリフェッチ・コード・エントリ・アドレス・テーブル503はコード・アドレス・アレイ503aとセレクト・アレイ503bとを備える。

【0076】ブリフェッチ作業用レジスタ・ファイル502は、ブリフェッチ・アドレス生成用プロセッサ501が、アドレス生成の作業用に用いるレジスタであり、実際には複数のレジスタから構成される。ただし、ブリフェッチ作業用レジスタ・ファイル502に何本のレジスタが含まれるかは実装時の決定となる。

【0077】ブリフェッチ作業用レジスタ・ファイル502に対する要件として、各レジスタのビット幅は、ブリフェッチするアドレスの計算が円滑に行なえるように、ブリフェッチ・アドレスのビット幅以上必要である。

【0078】ブリフェッチ・アドレス生成用プロセッサ501は、主プロセッサ100との間を、アドレス・バス121とブリフェッチ・コード・セレクト信号124とで信号の転送を行う。

【0079】また、ブリフェッチ・アドレス生成用プロセッサ501は、キャッシュ・アレイ400との間をブリフェッチ・アドレス・バス122で接続する。

【0080】これは、ブリフェッチ・アドレス生成用プロセッサ501が、タグ・アレイ401やフラグ・アレイ402の内容を参照／書き込みできるようにするためにある。どのアドレスのデータが現在キャッシュ・アレイ400に格納されていて、そのデータは有効か（読み書きしても問題ないか）という情報は、タグ・アレイ401やフラグ・アレイ402に含まれている。

【0081】ブリフェッチ・アドレス生成用プロセッサ501が、タグ・アレイ401やフラグ・アレイ402を参照し、ブリフェッチすべきデータがすでにキャッシュ・アレイ400にあるならば、ブリフェッチは行なわず、キャッシュ・アレイ400になれば、ブリフェッチ要求をフェッチ／リブレース制御部600に出すといったことを自律的に行なえるようになる。

【0082】ここでいう自律的とは、ブリフェッチアドレスの算出やブリフェッチ動作の可否の判断をブリフェッチ・コード・メモリ504に置かれているブリフェッチ・アドレス生成用プロセッサ501のプログラムの内容により決定されるという意味である。

【0083】ブリフェッチ／リブレース制御部600は、主プロセッサ100およびブリフェッチ・アドレス生成用プロセッサ501との間をデータ・バス123で

接続し、主メモリ300間をアドレス・バス312およびデータ・バス311で接続する。

【0084】さらに、プリフェッチ・アドレス生成用プロセッサ501との間を、キャッシュ・アクセス・アドレス・バス214およびプリフェッチ調停信号215の信号線で接続する。

【0085】プリフェッチ調停信号215は、フェッチ／リプレース制御部600とプリフェッチ・アドレス生成用プロセッサ501との間で処理を調停するためである。

【0086】調停の内容は、

(1) フェッチ／リプレース制御部600が動作中である場合、プリフェッチ調停信号215を通じて、プリフェッチ・アドレス生成用プロセッサ501に対して、プリフェッチ要求を受け付けられないことを通知する。

【0087】このとき、プリフェッチ・アドレス生成用プロセッサ501は、キャッシュ・アクセス・アドレス・バス214へのプリフェッチ・アドレスの送出を一時的に延期する。

(2) フェッチ／リプレース制御部600が、プリフェッチ要求を受け付け可能であるならば、プリフェッチ・アドレス生成用プロセッサ501に対して、プリフェッチ要求受け付け可能であることを、プリフェッチ調停信号215を通じて知らせる。

【0088】このとき、プリフェッチ・アドレス生成用プロセッサ501がプリフェッチ・アドレスの送出を延期している状態にあるならば、プリフェッチ・アドレス生成用プロセッサ501は延期していたプリフェッチ・アドレスの送出を再開する。

【0089】上述した構成に基づき、第1の実施形態の動作を説明する。

【0090】一般的に、キャッシュ・メモリは主メモリよりもメモリ容量が小さくてきているので、マイクロプロセッサが命令実行に際してキャッシュ・メモリをアクセスしたときに、必要なデータまたは命令がキャッシュ・メモリに格納されていないことがあることは充分にあり得る。

【0091】また、キャッシュ・メモリは、マイクロプロセッサが以前に実行したデータおよび命令を予め定める一定期間保持する。したがって、主プロセッサが最初にプログラムを開始したとき、あるいはアクセスしたときにはまだ未格納状態であり、キャッシュ・ミスとなり、主メモリへのアクセスが多発することになる。

【0092】その解決方法の1つとして、近々アクセスするであろうと思われるメモリラインを予測するとともに、主メモリにそのラインへのアクセスを行わせるプリフェッチ機構を対象とする。

【0093】まず、プリフェッチ・アドレス生成用プロセッサ501は、キャッシュ・アレイ400に対してプリフェッチすべきアドレスを生成する。

【0094】すなわち、プリフェッチ・アドレス生成用プロセッサ501は、予めプリフェッチ・コード・メモリ504に格納されているプリフェッチ・アドレス生成用プロセッサ501用の命令をプリフェッチ・コード・アドレス・バス512およびプリフェッチ・コード・データ・バス511を介して読み出し、読み出した命令を逐次実行することによって、プリフェッチ・アドレスを生成する。

【0095】生成されたプリフェッチ・アドレスを、キャッシュ・メモリ200内にあるフェッチ／リプレース制御部600に対してキャッシュ・アクセス・アドレス・バス214を介してプリフェッチ・アドレスを送出する。

【0096】プリフェッチ・アドレスを受けたフェッチ／リプレース制御部600は、受け取ったプリフェッチ・アドレスに対応したブロックのフェッチを行うとともに、主プロセッサ100が次に行うと予想されるアクセスに備え、待機する。

【0097】これにより、プリフェッチ・アドレスの決定をプログラムの実行時に行うことで、柔軟なプリフェッチ機構を実現できる。

【0098】すなわち、プリフェッチ・アドレス生成用プロセッサ501は、主プロセッサ100からの指示に従い、プリフェッチ・コード・メモリ504から命令を取出し、その命令を自律的に順次実行する。

【0099】なお、プリフェッチ・アドレス生成用プロセッサ501は、直接主メモリ300をアクセスすることではなく、必要なデータはすべてプリフェッチ・エンジン作業用レジスタ・ファイル502に格納する。

【0100】このプリフェッチ・アドレス生成用プロセッサ501の命令セット：アーキテクチャは、データ処理が十分高速に実行できるように水平型マイクロコードを模したコード体系から成る。

【0101】また、プリフェッチ・アドレス生成用プロセッサ501は、主プロセッサ100の指令によって主メモリ300に置かれたプリフェッチ・アドレス生成用プロセッサ501用の命令コード（以下、プリフェッチ・コードと称す）を、主メモリ300からアドレス・バス312およびデータ・バス311とフェッチ／リプレース制御部600とを介して読み出し、さらにプリフェッチ・コード・アドレス・バス512およびプリフェッチ・コード・データ・バス511を介して、プリフェッチ・コード・メモリ504にロードすることができる。

【0102】また、プリフェッチ・アドレス生成用プロセッサ501は、キャッシュ・アレイ400のタグ・アレイ401、フラグ・アレイ402、データ・アレイ403を、それぞれ対応するバス、タグ・データ・バス211、フラグ・データ・バス212、ブロック・データ・バス213とフェッチ／リプレース制御部600とを介して参照することができる。

【0103】そのとき、データ・アレイ403内の特定の範囲をアドレスと見なして、そのアドレスをブリフェッチする命令を備えている。

【0104】例えば、ある列のデータ・アレイ403に、1000、1004、0064、0000という4つのデータあるとする。上記命令は、このデータ・アレイにある4つのデータのうちの1つを取出す。

【0105】どのデータを取出すかはブリフェッチ・コード・メモリ504内のプログラムによって決定する。ここでは、2番目の1004というデータを取出したとすると、上記命令は、取出したデータ1004をアドレスとみなして、1004番地をブリフェッチするようにフェッチ/リプレース制御部600に指示を出す。

【0106】ここの例では、データ・アレイ403内の特定のデータに対してのみ処理をしているが、データ・アレイ内のデータを走査し、それぞれのデータが表すアドレスに対して順次ブリフェッチするという構成でもよい。

【0107】ブリフェッチ・エンジン500に対する主プロセッサ100からのブリフェッチの指示は、ブリフェッチすべきアドレスとブリフェッチの方法とをブリフェッチ・コード・セレクト信号124を与えられることによって行われる。

【0108】これらが与えられるとブリフェッチ・エンジン500は、ブリフェッチ・コード・セレクト信号124に対応づけられたブリフェッチ・コードのエントリ・ポイントを、ブリフェッチ・コード・エントリ・アドレス・テーブル503から読み出し、ブリフェッチ・アドレス生成用プロセッサ501に設定する。

【0109】ブリフェッチ・アドレス生成用プロセッサ501に対して、ブリフェッチ・コードの実行開始アドレスの設定を行なうのは、ブリフェッチ・アドレス生成用プロセッサ501自身である。

【0110】ブリフェッチ・コード・セレクト信号124がブリフェッチ・アドレス生成用プロセッサ501に入力されると同時に、セレクト・アレイ503bの検索が開始され、対応するエントリ・ポイントがコード・アドレス・アレイ503aから読み出される。この読み出された値をブリフェッチ・コードの開始アドレスとみなし、ブリフェッチ・コードをブリフェッチ・コード・メモリ504から取出す。

【0111】以上の処理は、ブリフェッチ・アドレス生成用プロセッサ501が自律的に行う。

【0112】それ以後、ブリフェッチ・アドレス生成用プロセッサ501は、設定されたブリフェッチ・コードの内容に従ってブリフェッチ用アドレスを順次に、フェッチ/リプレース制御部600に渡す。

【0113】ブリフェッチ・アドレス生成用プロセッサ501は、処理停止を意味するキャッシュ・コントロールを実行するか、主プロセッサ100から明示的にブリ

フェッチ処理の停止を指示されるまで自律的にアドレス生成処理を行う。

【0114】ブリフェッチ・コード・セレクト信号124に対応づけられたブリフェッチ・コードの生成は、プログラム言語を機械語に変換するコンパイラが、また必要に応じて機械語の実行履歴を解析するプロファイラとコンパイラが連係することで行われる。

【0115】コンパイラおよびプロファイラは、コードの特定の範囲（ループや関数など）ごとに解析を行い、これら範囲内において主プロセッサ100において、ブリフェッチの効果が局所最適となるようにブリフェッチ・コードを生成する。

【0116】局所最適とは、ある特定の状況下でのみ最大あるいは最大に近い効果が得られることであり、ここでは、特定の実行時間あるいは特定個所においてブリフェッチの効果が得られるような特別なブリフェッチ・コードを生成することをいう。

【0117】例えば、あるプログラムのある個所が複数実行されるものとして、奇数回目と偶数回目の実行の時にメモリ・アクセス・パターンが異なるものとする。

【0118】このとき、本方式では奇数回目のアクセス・パターンと偶数回目のアクセス・パターンに対して有効となるブリフェッチ・コードをそれぞれ用意して、実行時にそれぞれのブリフェッチ・コードをブリフェッチ・コード・メモリにロードしておいて、実行時にブリフェッチの効果がもっとも高くなると想定されるブリフェッチ・コードを選択して実行する。

【0119】主プロセッサ100およびブリフェッチ・エンジン500は並列して処理を行い、かつキャッシュ・アレイ400によって自律的にブリフェッチ処理を行うことが可能となる。

【0120】すなわち、ブリフェッチ・アドレス生成用プロセッサ501が主プロセッサ100と並列に動作することにより、ブリフェッチ・アドレス生成用プロセッサ501が主プロセッサ100とは独立にブリフェッチ・アドレスを生成する。

【0121】生成されたブリフェッチ・アドレスは、フェッチ/リプレース制御部600に渡され、フェッチ/リプレース制御部600が主メモリ300にあるブリフェッチ・アドレスの内容をキャッシュ・アレイ400に格納することになる。

【0122】したがって、本実施形態により、柔軟なブリフェッチ処理および主プロセッサ100の処理能力に影響を与えることなくブリフェッチを行うことができる。

【0123】次に、ブリフェッチ・アドレスの決定をプログラムの実行時に行う方法を説明する。

【0124】まず、ブリフェッチを行う場合の方法を説明する。ブリフェッチを行う場合の処理方法を説明するためのフローチャートを示した図2を参照すると、方法

の処理ステップS11において、主プロセッサ100がメモリアクセスをキャッシュ・メモリ200に対して要求する。アクセス要求はアドレス・バス121とプリフェッチ・コード・セレクト信号124を通じてプリフェッチ・アドレス生成用プロセッサ501に伝えられる。

【0125】このプリフェッチ・コード・セレクト信号124は、あらかじめプリフェッチ・コード・メモリ504に書き込まれて保持されているプリフェッチ・アドレス生成用プロセッサ501が実行すべき命令列を、主プロセッサ100が指定するために使われる。

【0126】処理ステップS12において、プリフェッチ・アドレス生成用プロセッサ501は、主プロセッサ100からプリフェッチ・コード・セレクト信号124を受け取ると、プリフェッチ・コード・エントリ・アドレス・テーブル503を用いて、プリフェッチ・コード・メモリ504にある命令列の先頭アドレスを検索する。

【0127】プリフェッチ・コード・エントリ・アドレス・テーブル503は、プリフェッチ・コード・セレクト信号124を通してプリフェッチ・アドレス生成用プロセッサ501に伝えられるプリフェッチ指示コードと、それに対応するプリフェッチ・コード・メモリ504内にある命令列の先頭アドレスを一对の組として保持している。そのため、プリフェッチ・コード・セレクト信号124が与えられると、それに対応する命令列の先頭アドレスをプリフェッチ・アドレス生成用プロセッサ501に返す。

【0128】処理ステップS13において、プリフェッチ・アドレス生成用プロセッサ501は、処理ステップS12の検索で得られた命令から命令の実行を開始する。

【0129】処理ステップS14において、プリフェッチ・アドレス生成用プロセッサ501が、プリフェッチ・アドレスを生成する命令を実行した場合は、プリフェッチ・アドレス生成用プロセッサ501は、フェッチ／リプレース制御部600に対して、キャッシュ・アクセス・アドレス・バス214を通じてプリフェッチすべきアドレスを送出する。

【0130】処理ステップS15において、プリフェッチ・アドレス生成用プロセッサ501は、受け取ったアドレスに対応するキャッシュ・ブロックを入れ替える。

【0131】キャッシュ・ブロックの入れ替えは、フェッチ／リプレース制御部600の役割である。プリフェッチ・アドレス生成用プロセッサ501はプリフェッチ・アドレスの生成のみを行う。そのときのキャッシュ・ブロック入れ替えの経路は次のようになる。

(1) プリフェッチ・アドレス生成用プロセッサ501が入れ替え対象となるアドレスを生成する。

(2) 入れ替え対象のアドレスをキャッシュ・アクセス・アドレス・バス214を通じてフェッチ／リプレース

制御部600へ転送する。

(3) フェッチ／リプレース制御部600が入れ替え処理を行なう。

【0132】処理ステップS16において、プリフェッチの停止は、プリフェッチ・コードの実行が終了するか、あるいは主プロセッサ100がプリフェッチ・コード・セレクト信号124を通じてプリフェッチの停止を指示するまで行われる。

【0133】次に、プリフェッチを行わない場合の方法を説明する。プリフェッチを行わない場合の処理方法のフローチャートを示した図3を参照すると、処理ステップS21において、主プロセッサ100がメモリアクセスをキャッシュ・メモリ200に対して要求する。

【0134】プリフェッチを行わない場合は、プリフェッチ・コード・セレクト信号124を通じてプリフェッチ・アドレス生成用プロセッサ501に対して、プリフェッチを行わないように指示する。

【0135】処理ステップS22において、プリフェッチ・アドレス生成用プロセッサ501は、プリフェッチを行わず、単にキャッシュ・アレイ400のアクセスを行う。

【0136】次に、プリフェッチ・コードの読み出しおよびプリフェッチ・コードの入れ替え方法のフローチャートを示した図4を参照すると、処理ステップS31において、主プロセッサ100は、プリフェッチ・アドレス生成用プロセッサ501に対して、プリフェッチ・コード・セレクト信号124を通じてプリフェッチ・コード・メモリの入れ替えを指示する。

【0137】処理ステップS32において、プリフェッチ・アドレス生成用プロセッサ501は、アドレス・バス121から受け取ったアドレスで主メモリ300からプリフェッチ・コードを読み出す。

【0138】処理ステップS33において、主プロセッサ100は、読み出したサイズと読み出したプリフェッチ・コードに対応するプリフェッチ指示コードを、データ・バスを通じてプリフェッチ・アドレス生成用プロセッサ501に与える。

【0139】プリフェッチ・コードはプリフェッチ・アドレス生成用プロセッサ501で実行される命令列のことをいう。このプリフェッチ・コードは、主プロセッサ100のメモリアクセス・パターンに応じて複数種類がプリフェッチ・コード・メモリ504に格納される。

【0140】一方、プリフェッチ指示コードは、複数あるプリフェッチ・コードのそれぞれに対応づけられた識別符号をさしている。具体的には、プリフェッチ指示コードは、プリフェッチ・コードに一对一に与えられた整数値、あるいはビット・パターンで、それはセクタ・アレイ503bに格納される。

【0141】コード・アドレス・アレイ503aには、プリフェッチ指示コードに対応したプリフェッチ・コー

10

20

30

40

50

ド開始アドレスが格納される。

【0142】処理ステップS34において、プリフェッチ・アドレス生成用プロセッサ501は、主メモリ300から読み込んだプリフェッチ・コードをプリフェッチ・コード・メモリ504にロードする。

【0143】同時に、ロードした命令列の先頭アドレスと主プロセッサ100から与えられたプリフェッチ指示コードをプリフェッチ・コード・エントリ・アドレス・テーブル503に登録する。

【0144】次に、コンパイラのコード生成方法を説明\*10すと、

```
struct foo {
    struct foo * next;
    ...
}

struct foo * ptr;
/* ptr->nextの指すアドレスをプリフェッチする/*
#pragma prefetch (prefetch addr=ptr->
next)
for (ptr=list head; ptr !=0; ptr=ptr->
next) {
    .....
}
```

【0146】また、このとき主プロセッサ100用の暫定的な命令列を生成する。

【0147】処理ステップS43において、コンパイラは、推測したプリフェッチ・パターンをプリフェッチ・アドレス生成用プロセッサ501が出力するようなプリフェッチ・コードを生成する。

【0148】処理ステップS44において、プリフェッチ・パターンが複数推測される場合は、それぞれについてプリフェッチ・コードを生成する。

【0149】処理ステップS45において、生成したプリフェッチ・コードに対して、生成プログラム中で一意となるようなプリフェッチ指示コードを定義する。

【0150】処理ステップS46において、コンパイラは、処理ステップS42で生成した主プロセッサ100用の命令のうち、プリフェッチ・コードを利用するメモリアクセス命令をメモリアクセスと共にプリフェッチを行う命令に置き換える。

【0151】置き換える際には、処理ステップS44で生成したプリフェッチ・コードのうち、もっともプリフェッチの効果が高いと予測されるプリフェッチ・コードに対応したプリフェッチ指示コードを、プリフェッチ・アドレス生成用プロセッサ501に与えるように、メモリアクセス命令を置き換える。

【0152】上述したように、第1の実施形態では、プリフェッチのパターンを複数設定でき、かつそれらを選択できることにより、従来のプリフェッチが有効でなかったプログラムに対してもプリフェッチを行うことができ、主メモリがデータを要求してから使用できるまでの

\* するためのフローチャートを示した図5を参照すると、処理ステップS41において、コンパイラは、ループやリンク・ポインタの追跡処理などプリフェッチが有効であるコードを認識すると、メモリアクセス・パターンを解析し、それについて有効なプリフェッチ・パターンを推測する。

【0145】また、プリフェッチ・パターンの解析は、ユーザがソース・コード中に埋め込むコンパイラ・ディレクティブによって指定できる。例えばC言語での例を示

平均的なメモリアクセスのレイテンシを隠蔽する効果がある。

【0153】プリフェッチ・コード・メモリ504に格納されるコードのみを入れ替えることで、主プロセッサ100に対するコードを変更することなくプリフェッチ・パターンを変更することができる。

【0154】このため、プログラムを再コンパイルすることなくプログラムのチューニングを行うことができる。

【0155】上述した各プログラムと各プロセッサとの関係は、キャッシュ・メモリ装置に対してどのようなプリフェッチを行なうのかを指示するのは主プロセッサ100で実行するプログラムであり、具体的なプリフェッチの挙動を指示するのがプリフェッチ・コードになる。両プログラムがそろってはじめて本発明のキャッシュ・メモリ装置の効果が得られる。

【0156】本発明の第2の実施形態を説明する。第2の実施形態の主要部をブロック図で示した図6を参照すると、この第2の実施形態が第1の実施形態と異なる点は、プリフェッチ・コード・メモリ504に代えてプリフェッチ・コード領域を主メモリ300に設けていることである。それ以外の構成要素は第1の実施形態と同様であるから、ここでの構成の説明は省略する。

【0157】なお、図6では図1との相違点の理解を容易にするために、主メモリ300内にプリフェッチ・コードを示す領域を仮想的に表示してある。

【0158】すなわち、プリフェッチ・コードは主メモリ300に格納されるが、主メモリ300にはプリフェ

ッチ・コード専用となる領域は物理的には存在せず、主プロセッサ100と専用プロセッサ501とで主メモリ300を共用する。

【0159】この場合、ハードウェアは簡素化されるが同一の主メモリ300に対して主プロセッサ100とブリフエッチ・アドレス生成用プロセッサ501の両方からメモリフェッチが行われるため、主メモリ300からの命令供給能力に特に問題のない処理量のシステムに対して適用することができる。

【0160】本発明の第3の実施形態を説明する。第3の実施例の主要部をブロック図で示した図7を参照すると、この第3の実施形態が第1の実施形態と異なる点は、ブリフエッチ・コード・メモリ504に代えて、ブリフエッチ・コードをキャッシュ・アレイ400のデータ・アレイ403に格納していることである。それ以外の構成要素は第1の実施形態と同様であるからここの構成の説明は省略する。

【0161】なお、図7では図1との相違点の理解を容易にするために、キャッシュ・アレイ400のデータ・アレイ403内にブリフエッチ・コードを示す領域を仮想的に表示してある。

【0162】すなわち、ブリフエッチ・コードは主メモリ300に格納されるが、主メモリ300にはブリフエッチ・コード専用となる領域は物理的には存在せず、キャッシュ・アレイ400にブリフエッチ・コードを置く場合、キャッシュ・アレイ400内に物理的にブリフエッチ・コード用のメモリを設けるのではなく、すでにキャッシュ・アレイ400にあるデータ・アレイ403にデータを格納するということである。

【0163】ブリフエッチ・コードは、キャッシュ・アレイ400のデータ・アレイ403を間借していることになる(実際には、キャッシュ・アレイ400は主プロセッサ100が使うコード/データか、あるいはブリフエッチ・アドレス生成用プロセッサ501のブリフエッチ・コードかを区別することなく、指定されたものを内部に取り込む)。

【0164】また、キャッシュ・メモリ200のデータ・アレイ403にブリフエッチ・コードを置くということは、データ・アレイ403だけでなくタグ・アレイ401とフラグ・アレイ402にも「キャッシュ・メモリのこの領域は、ブリフエッチ・コードを格納している」「ブリフエッチ・コードの格納領域は勝手に入れ替えてはいけない」等の情報も格納しておく必要がある。

【0165】したがって、データ・アレイ403、フラグ・アレイ402、タグ・アレイ401、の3者の協働が必要ということであり、コードそのものがデータ・アレイ403に配置されて、タグ・アレイ401、フラグ・アレイ402には管理情報が格納されるということである。

【0166】この場合、ハードウェアは簡素化されるが

同一のキャッシュ・アレイ400に対して主プロセッサ100とブリフエッチ・アドレス生成用プロセッサ501の両方からメモリフェッチが行われるため、キャッシュ・メモリ200からの命令供給能力に特に問題のない処理量のシステムに対して適用することができる。

【0167】次に、本発明の第4の実施形態を説明する。第4の実施例の主要部をブロック図で示した図8を参照すると、第1の実施形態との相違点は、ブリフエッチ・コード・セレクト信号124のビット幅を拡張していることである。それ以外の構成要素は第1の実施形態と同様であるから、ここの構成の説明は省略する。

【0168】メモリアクセス命令はブリフエッチすべきアドレスを「陽に」渡せるようにブリフエッチ・コード・セレクト信号124のビット幅を大きくする条件の下に、メモリアクセス命令にブリフエッチ・コードのアドレスを指定できるだけの情報を備える。

【0169】この実施形態では、メモリアクセス命令の中にブリフエッチ指示コードとともに関連するアドレスを埋め込み、これらをブリフエッチ指示コードとして専用プロセッサ501に送る。この方法により、従来からあるブリフエッチ専用命令と同等の機能を実現できる。

【0170】上述した情報を備えることで、主プロセッサ100がブリフエッチ・アドレス生成用プロセッサ501に対してブリフエッチ指示コードを指示する。

【0171】ここの「陽に」とは、Explicitの意味で、従来のプロセッサの場合だと、プロセッサの持つ命令の中に、指定されたアドレスのブリフエッチを行なう、という命令があり、その命令を使うことでブリフエッチを行うようになっていたが、ここでは、プロセッサがブリフエッチ・アドレスを明確に指定する、という意味で「陽に」という言葉を使っている。

【0172】その指示を受けて、ブリフエッチ・アドレス生成用プロセッサ501がブリフエッチ・コード・エントリ・アドレス・テーブル503からブリフエッチ・コードの先頭アドレスを検索する代わりに、主プロセッサ100が直接ブリフエッチ・コードの先頭アドレスを渡すようにしてある。

【0173】この第4の実施形態の場合、実行開始を本発明の方式に比べてブリフエッチ・コード・セレクト信号124が入力されてから、ブリフエッチ・アドレス生成用プロセッサ501がブリフエッチ・コードの実行を開始するまでの時間を短くすることができる。

【0174】

【発明の効果】上述したように、本発明のソフトウェア制御可能なキャッシュ・メモリ装置、制御方法およびプログラムは、主プロセッサおよび主メモリ間に設けられるキャッシュ・メモリが、プログラム実行時に、ブリフエッチ・アドレスをソフトウェア制御で決定するブリフエッチ手段として、実行すべきブリフエッチ・コードの指定を主プロセッサから指示されるとともにブリフエ



チ・コードに対応した命令の実行によりプリフェッチ・アドレスを生成するプリフェッチ・アドレス生成用プロセッサを含み、かつ主プロセッサと並列処理を行うプリフェッチ・エンジンを備えるので、プリフェッチのパターンを複数設定でき、かつそれらを選択できることにより、従来のプリフェッチが有効でなかったプログラムに対してもプリフェッチを行うことができ、主メモリがデータを要求してから使用できるまでの平均的なメモリアクセスのレイテンシを隠蔽する効果がある。

【0175】プリフェッチ・コード・メモリに格納されるコードのみを入れ替えることで、主プロセッサに対するコードを変更することなくプリフェッチ・パターンを変更することができる。

【0176】このため、プログラムを再コンパイルすることなくプログラムのチューニングを行うことができる。

【図面の簡単な説明】

【図1】本発明の第1の実施形態の主要構成を示すブロック図である。

【図2】本発明におけるプリフェッチを行う場合の処理方法を説明するためのフローチャートである。

【図3】プリフェッチを行わない場合の処理方法を説明するためのフローチャートである。

【図4】プリフェッチ・コードの読み出しおよびプリフェッチ・コードの入れ替え方法を説明するためのフローチャートである。

【図5】コンパイラのコード生成方法を説明するためのフローチャートである。

【図6】本発明の第2の実施例の主要部を示すブロック図である。

【図7】本発明の第3の実施例の主要部を示すブロック図である。

【図8】本発明の第4の実施例の主要部を示すブロック\*

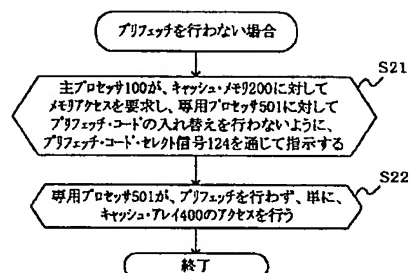
\*図である。

【図9】従来のプリフェッチ機構を有する方式の一例を示すブロック図である。

【符号の説明】

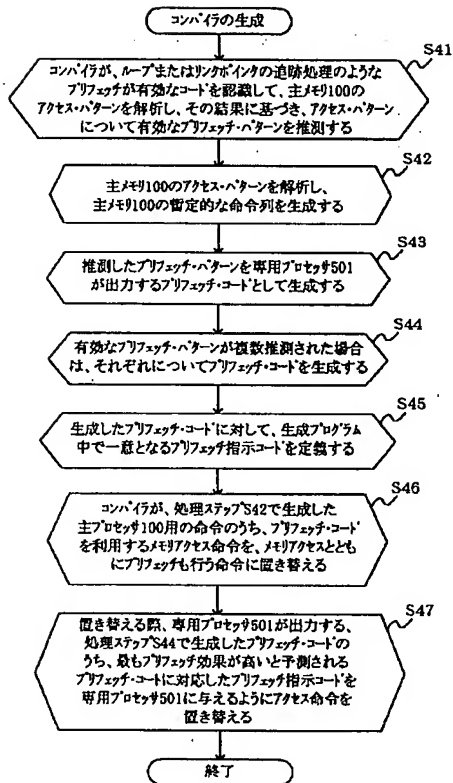
100	主プロセッサ
121	アドレス・バス
122	プリフェッチ・アドレス・バス
123	データ・バス
124	プリフェッチ・コード・セレクト信号
200	キャッシュ・メモリ
211	タグ・データ・バス
212	フラグ・データ・バス
213	ブロック・データ・バス
214	キャッシュ・アクセス・アドレス・バス
215	プリフェッチ調停信号
300	主メモリ
400	キャッシュ・アレイ
401	タグ・アレイ
402	フラグ・アレイ
403	データ・アレイ
500	プリフェッチ・エンジン
501	プリフェッチ・アドレス生成用プロセッサ
502	プリフェッチ・エンジン作業用レジスタ・ファイル
503	プリフェッチ・コード・エントリ・アドレス・テーブル
503a	コード・アドレス・アレイ
503b	セクタ・アレイ
504	プリフェッチ・コード・メモリ
511	プリフェッチ・コード・データ・バス
512	プリフェッチ・コード・アドレス・バス
600	プリフェッチ/リブレース制御部

【図3】





【図5】



【図6】

